Linguistic structure is an evolutionary trade-off between simplicity and expressivity

Kenny Smith (kenny@ling.ed.ac.uk), Monica Tamariz & Simon Kirby
Language Evolution and Computation Research Unit, School of Philosophy, Psychology & Language Sciences,
University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, UK

Abstract

Language exhibits structure: a species-unique system for expressing complex meanings using complex forms. We present a review of modelling and experimental literature on the evolution of structure which suggests that structure is a cultural adaptation in response to pressure for expressivity (arising during communication) and compressibility (arising during learning), and test this hypothesis using a new Bayesian iterated learning model. We conclude that linguistic structure can and should be explained as a consequence of cultural evolution in response to these two pressures.

Keywords: language; structure; cultural evolution; learning; communication

Introduction

Human language is unique among the communication systems of the natural world in that it is exhibits a rich combinatorial and compositional structure: language provides a generative system for productively combining meaningless elements (e.g. speech sounds) to form meaning-bearing units (morphemes), which are further recombined to yield complex units (phrases) whose meaning is derived in a predictable manner from the meaning of their component parts and their manner of composition. This allows us massive expressive potential: at least at a first approximation, anything you can think you can express in language. No other species has a communication system providing anything approaching this expressive power: why do humans?

One explanation for the presence of *structure*[1] in human language appeals to biological evolution under natural selection (Pinker & Bloom, 1990): language is fundamentally a biological trait, being underpinned by some innate language-specific apparatus; the ability to communication propositions which a structured language provides is adaptive, since it facilitates social interaction and ultimately increases fitness; therefore, structure in language represents a biological adaptation to facilitate communication. A second account explains structure in language as a consequence of *cultural*, rather than biological, evolution (Christiansen & Chater, 2008). Rather than language structure reflecting an evolved domain-specific learning apparatus, the idea is that languages have adapted over repeated episodes of learning and production (a process sometimes called *iterated learning*) in response to weaker, domain-general constraints arising from the biases of language learners. We have previously termed this evolutionary process *cultural selection for learnability* (Brighton, Kirby, & Smith, 2005). A range of models and experiments show that cultural selection for learnability leads to the evolution of structure, under certain assumptions about the nature of transmission and the biases of language learners. Under a strong interpretation of this account, language's function for communication could be seen as an epiphenomenon: structured language provides a powerful medium for communication, but language structure is not 'for' communication.

Here we present a new model of iterated learning, motivated by recent experimental work, which goes some way to reconcile these two viewpoints. We draw from the biological account the insight that the alignment between language's apparent function as a system for expressing propositions and its structure, tailor-made for just such a purpose, is unlikely to be a fortuitous coincidence. We draw from cultural evolutionary account two insights: 1) biological evolution is not the only evolutionary process which might act to shape language: cultural evolution, a necessary consequence of the fact that language is socially learned, is a second such mechanism; 2) selection for learnability will impact on language during its transmission. This model suggests that structure arises from cultural evolution when language is under pressure to be expressive *and* learnable: pressure for expressivity arises from language use in communication, language learning by naive individuals introduces a pressure for simplicity arising from domain-general preferences for compressibility in learning. Crucially, both must be in play: pressure for expressivity or simplicity alone does not lead to structure. Structure in language is a linguistic adaptation, not a biological adaptation, in response to competing pressures for expressivity and learnability (Kirby, Cornish, & Smith, 2008; Steels, 2012).

Cultural evolution of structure: previous work

Models of the cultural evolution of linguistic structure typically emphasise the role of learnability constraints in driving the evolution of compositionality. Specifically, language is under pressure to be *compressible*: to allow the formation of compressed mental representations, i.e. simple grammars. This pressure for compressibility is inherent in learning (Chater & Vitanyi, 2003), and can be amplified by other constraints acting on language transmission (e.g. the mismatch between the infinite expressivity of languages and the finite set of data from which such languages must be learned). Learning and transmission therefore favour languages which admit to compressed representations, i.e. which permit generalisations. Recursive compositionality is one such generalisation (e.g. Steels, 1998; Kirby, 2002; Brighton, Smith, & Kirby, 2005), and therefore represents an adaptation by lan-

[1] We use the term structure as a shorthand for combinatoriality and/or compositionality.

guage in response to pressures inherent in transmission and learning. However, compressibility is not the only constraint on learnability in these models: they typically include some learner bias in favour of languages which embody a one-to-one mapping between meaning and form (which happen to be communicatively functional mappings), e.g. by implementing indirect (Kirby, 2002) or direct (Steels, 1998) competition between meanings which map to a single form. Brighton, Smith, and Kirby (2005) show that, if this bias against 1-to-1 mappings is absent, the pressure for compressibility acting in isolation leads to *degenerate*, not structured, languages, where all meanings map to a single maximally-ambiguous form. While this might suggest that such a 1-to-1 bias would be adaptive, Smith (2004) shows that such a bias is unlikely to evolve for its (eventual) communicative payoff, and concludes that the one-to-one bias must be a product of domain-general cognitive biases. Again, this suggests that the utility of language for communication might be a side-effect of learnability pressures alone.

Diffusion-chain experiments with adult human participants have also been used to investigate the impact of cultural selection for learnability. Kirby et al. (2008) report two experiments in which participants are trained on a miniature language which provides labels for objects (coloured moving shapes, e.g. a red square bouncing), and are then prompted to produce labels for a further set of objects. Participants are organised into a diffusion chain, such that the labels produced by the nth participant in a given chain provide the training data for participant $n+1$ in that chain. The first participant in each chain is trained on a unstructured *holistic* system, where each object is associated with a unique random label (and therefore shared elements of meaning do not map to shared components of form).

Across two experiments, Kirby et al. (2008) show that languages change as a result of their transmission to be more learnable: the languages produced later in a chain of transmission are learnt with greater accuracy. In their Experiment 1, this is achieved by the languages becoming *simple*: the languages lose distinctions. In the most extreme case, this results in a degenerate language in which all objects (with one exception) are associated with a single, highly-ambiguous label. Simplification facilitates learning at the expense of expressivity: while the initial holistic languages have high expressive potential, the languages which ultimately emerge allow only a few contrasts between objects to be signalled linguistically. However, there is no pressure for expressivity in this experiment: the language is under pressure to be learnable, but given the lack of a communicative task, under very little pressure to provide distinct labels for distinct objects.

In their Experiment 2, an artificial pressure for expressivity was introduced: homonyms (labels paired with multiple objects) were eliminated during the process of sampling from the nth participant's productions to yield the training data for participant $n+1$. As in Experiment 1, the languages became more learnable, but this was achieved by the development of compositional structure: colour, shape and motion came to be encoded in separate 'morphemes' of a multi-morphemic words. These structured languages are both learnable and expressive, allowing all distinctions between objects to be encoded linguistically.

Garrod, Fay, Lee, Oberlander, and MacLeod (2007) present a task in which participants are required to communicate a set of pre-specified concepts using drawings. Participants who repeatedly play the game together develop an expressive system of symbol-like graphical representations to communicate these concepts. This system of communication is holistic: each symbol is an idiosyncratic, stand-alone entity. Theisen-White, Kirby, and Oberlander (2011) present a modified version of this paradigm, integrating the dyadic context for communication with the diffusion-chain method from Kirby et al. (2008). An initial pair play a variant of the communication game from Garrod et al., using a modified set of concepts designed to provide a basis for systematic structure (e.g., teacher, school, teaching; firefighter, fire station, fire-fighting). The drawings produced by that pair during communication are then observed by a fresh pair of participants, who go on to communicate together, and so on. The system of communication is therefore under pressure to be both expressive (communicatively functional) *and* learnable (by the naive individuals during the observation phase). Theisen-White et al. find that the sets of drawings become more structured over these chains of transmission: the drawings develop component parts which refer to the domain (e.g. teaching, fire-fighting) and the category (e.g. person, building, activity).

These experimental results are therefore consistent with the modelling literature reviewed above and suggest a three-way contrast: pressure for compressibility alone results in degenerate languages (Kirby et al., 2008, Experiment 1); pressure for expressivity but not learnability (Garrod et al., 2007) leads to holistic systems; pressure for expressivity (from artificial filtering or, better, communication) leads to structure (Kirby et al., 2008; Theisen-White et al., 2011). However, no one model or experimental paradigm completely decouples learnability and expressivity: below, we present a model which does this, and which demonstrates this link between expressivity, learnablity and structure more conclusively.

The model

We model individuals as rational learners who infer a distribution over possible languages (meaning-form mappings), and use those languages to communicate. Learners have a (parameterised) prior preference for simple, compressible languages, and during interaction a (parameterised) tendency to avoid utterances which are ambiguous in context.

Model of languages

A language consists of a system for expressing meanings using forms. We consider the simplest possible meanings and forms which are nonetheless capable of evidencing systematic structure: meanings are vectors of length v, where each

element in the vector takes one of w possible values. Similarly, forms are of length l, where each character is drawn from some alphabet Σ. We take $v = w = l = |\Sigma| = 2$, which yields a set of meanings $\mathcal{M} = \{00, 01, 10, 11\}$ and a set of forms $\mathcal{F} = \{aa, ab, ba, bb\}$. This gives a space of 256 possible languages, including degenerate, compositional and holistic mappings: see Table 1 for examples.[2]

Hypotheses

Learners infer a distribution over languages: the space of hypotheses is therefore the space of possible distributions over all 256 languages. Following Burkett and Griffiths (2010), we use a Dirichlet process prior (Ferguson, 1973), characterised by concentration parameter α and base distribution G_0. The parameter α determines how many languages feature in this distribution: low alpha (we use $\alpha = 0.1$) corresponds to an *a priori* belief that the majority of the probability mass will be on a single language. The base distribution is a distribution over languages, and would be the prior if learners only considered single-language hypotheses.

Our base distribution encodes a preference for simplicity, operationalised as a preference for languages whose description is compressible. Intuitively, degenerate languages permit more compressed descriptions than compositional languages; holistic languages are, by definition, incompressible. The prior used in Kirby, Dowman, and Griffiths (2007) captures this intuition: it assigns higher probability to languages in which fewer forms are used to convey a given set of meanings. We simply apply this metric both over the full set of meanings and specific feature values (see Appendix). This prior splits the space of 256 possible languages into 12 language classes, based on the number of forms in the language and the regularity with which feature of meaning are mapped to components of form: the priors for individual languages are depicted in Fig. 1, with example languages from some pertinent classes in Table 1. As the figure shows, the prior yields the desired ranking of languages: more compressible languages (i.e. with fewer forms) are preferred, but within those languages with a given number of forms, there is a preference for languages which consistently map feature values in the meaning to a single character in the corresponding position in the form.

Likelihood

We sample a form f from the distribution $P(f|h,C,t)$, which specifies the probability of f given hypothesis h, a *context* of utterance (a set of meanings) C, and *topic* $t \in C$, which the speaker attempts to discriminate from the other meanings in C. $P(f|h,C,t) = P(l|h) \cdot P(f|l,C,t)$: we simply sample a language l from the speaker's hypothesis, then given that language and the context, sample an utterance. We include a parameterisable preference to avoid ambiguity during this latter step, following the model of pragmatics provided by Frank and Goodman (2012). Assuming some small probability of error on production ε:

$$P(f|l,C,t) \propto \begin{cases} \left(\frac{1}{a}\right)^{\gamma}(1-\varepsilon) & \text{if } t \text{ is mapped to } f \text{ in } l \\ \frac{\varepsilon}{|\mathcal{S}|-1} & \text{if } t \text{ is not mapped to } f \text{ in } l \end{cases},$$

where we normalise over all possible forms from $\mathcal{F}$. a is ambiguity, the number of meanings in C that map to form f in l, and γ specifies the extent to which utterances which are ambiguous in context are penalised. If $a = 1$ (f is unambiguous in this context) and/or $\gamma = 0$ then this yields a model of production where the 'correct' form is produced with probability $1-\varepsilon$. However, when $\gamma > 0$ and f is ambiguous in context (i.e. $a > 1$), then the 'correct' mapping from t to f is less likely to be produced, and the remaining probability mass is spread equally over the other possible forms. Therefore, $\gamma > 0$ introduces a penalty for languages whose utterances are ambiguous in context. We use $\varepsilon = 0.05$, $|C| = 3$, and vary γ.

Inference

Exact inference over this hypothesis space in intractable: instead, following Burkett and Griffiths (2010), we use a Gibbs sampler based on the Chinese Restaurant Process to sample

Table 1: Example languages from three important classes.

	Form		
Meaning	degenerate	compositional	holistic
00	aa	aa	aa
01	aa	ba	ab
10	aa	ab	ba
11	aa	bb	bb

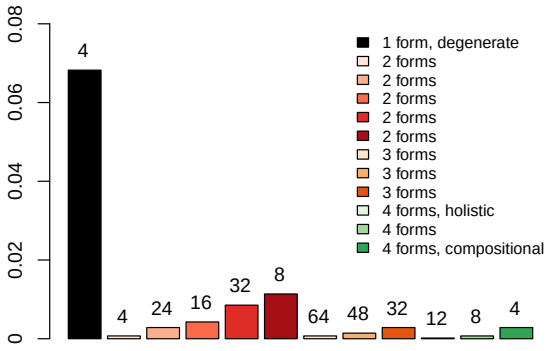

Figure 1: Probability in G_0 for individual languages in each class, arranged by number of distinct forms, and (within a given number of forms) increasing compressibility. Annotations give the number of languages per class, all of which have equal prior probability.

[2] We assume that the first meaning feature is expressed in the first form character. Without this constraint, it is impossible to specify holistic languages given this small form space: e.g. the holistic language in Table 1 is compositional if we allow the first meaning feature to map to the second form character. It would be possible to distinguish between holistic and compositional systems without this constraint given $|\Sigma| > 2$, but to minimise runtimes we opted for $|\Sigma| = 2$ and a constrained definition of compositionality.

a hypothesis direct from the posterior. As described below, learners acquire an expanding set of observed utterances during their lifetime: we run the inference over the most recent $r = 80$ observations, in order to improve simulation runtimes.

Transmission in populations

Following the experimental methods employed by Theisen-White et al. (2011) and Garrod et al. (2007), we compare two types of population: in *chains*, simulated agents are organised into pairs, are trained on data produced by the previous pair (see below), and then interact, producing data which the next generation in the chain (a new, naive pair of simulated individuals) are trained on. In *dyads* exactly the same regime of training and interaction is observed. However, naive individuals are not introduced at each generation: rather, the same individuals are trained on their own productions from the previous phase of interaction.[3] The contrast between chains and dyads allows us to manipulate the pressure for learnability: in chains, where naive individuals are introduced at every generation, the pressure for learnability (i.e. the influence of the prior preference for simplicity) is likely to be relatively strong. In dyads, in contrast, there is only one episode of transmission to naive individuals (at generation 1), and consequently the pressure for simplicity arising from the prior is substantially diminished.

Training During training, the pair are presented with a shared set of 20 form-meaning pairs, produced by the preceding pair during interaction or (for the first generation only) a shared set of 20 form-meaning pairs generated from a randomly-selected fully-expressive holistic language (this initialisation with holistic languages is inspired by the experimental work discussed above). This data is added to each agent's memory (which will be empty for individuals in chains), and then a hypothesis is sampled from the posterior.

Interaction After training, the pair interact for 40 rounds. At each round of interaction, one individual acts as teacher and the other as learner. The teacher is prompted with a randomly-selected context and topic, and samples a form from their hypothesis. The learner adds the observed form-meaning-context triple to its memory, and samples an updated hypothesis. The roles of teacher and learner then switch, and a new round is played.

Transmission The 20 form-meaning pairs produced by one randomly-selected member of the pair at generation n is used as the training data for the pair at generation $n+1$. In accordance with an ongoing set of human experiments based on these models, the context is stripped from these observations: in other words they contain only form-meaning pairs, with a context consisting solely of the target.

[3] Training dyads on their own productions ensures that the configuration of the model is identical for dyads and chains. We ran an additional set of dyad simulations with a modified transmission regime, such that pairs are trained on the initial target language and go on to interact repeatedly but are not retrained on their productions from the last round of interaction (i.e. there is no training phase after generation 1): this produces results which are highly similar to dyads with transmission at every generation, showing that the retraining step does not introduce some additional conservative tendency.

Results

The results (Fig. 2) match the predictions of our hypothesis, and are consistent with the experimental results described above. When there is pressure for learnability arising from transmission to naive individuals, but no pressure for expressivity (achieved by using the chain population model and setting γ for interaction to 0), the final distribution is dominated by degenerate languages, as in Kirby et al. (2008), Experiment 1. Note that this preference for degenerate languages is even stronger than that seen in the prior: given the parameters of the model, in particular the low concentration parameter for the Dirichlet process prior, this exaggeration of the prior is as predicted by Burkett and Griffiths (2010). In contrast, in the condition where there is expressivity pressure but little pressure for learnability (dyads, $\gamma = 3$), the initial holistic languages, (expressive but not compressible) persist. Members of the dyad constantly replenish their own evidence that the language is holistic: consequently, the initial holistic language is locked in. This matches the experimental results obtained for dyads (Garrod et al., 2007): due to the lack of transmission to new individuals, there is little pressure for compressibility to counteract lock-in and expressivity requirements during interaction, and structure does not emerge. Note also that this result holds despite the fact that we set a fairly low memory limit for individuals ($r = 80$). Finally, when there is pressure for both learnability and expressivity (chains, $\gamma = 3$), we see structured languages emerge: the final distribution is dominated by *a priori* unlikely expressive languages, but among these it is the *a priori* most likely languages, the compositional languages, that dominate. Again, this matches our hypothesis and the experimental data from Theisen-White et al. (2011).

Discussion

Our model shows that pressure for expressivity or simplicity alone does not lead to the emergence of structure: only when both pressures are at play does structure emerge. Furthermore, *only* cultural evolution is required for structure: we can explain why language is structured without recourse to invoking an evolved, domain-specific faculty of language.

As well as corresponding closely with existing modelling and experimental data, these findings make sense of the distribution of structure in the communication systems of non-human animals. Many small but expressive communication systems exist in nature, a classic example being alarm calling systems, which allow the discrimination of several referents (predators), but do so using vocalisations which are holistic and unlearned (Fitch, 2000). Learned vocal communication systems are witnessed in many species of bird, as well as being patchily distributed among mammals (Fitch, 2000): strikingly, song, the classic example of (combinatorial, not compositional) structure in animal communication, occurs in pre-

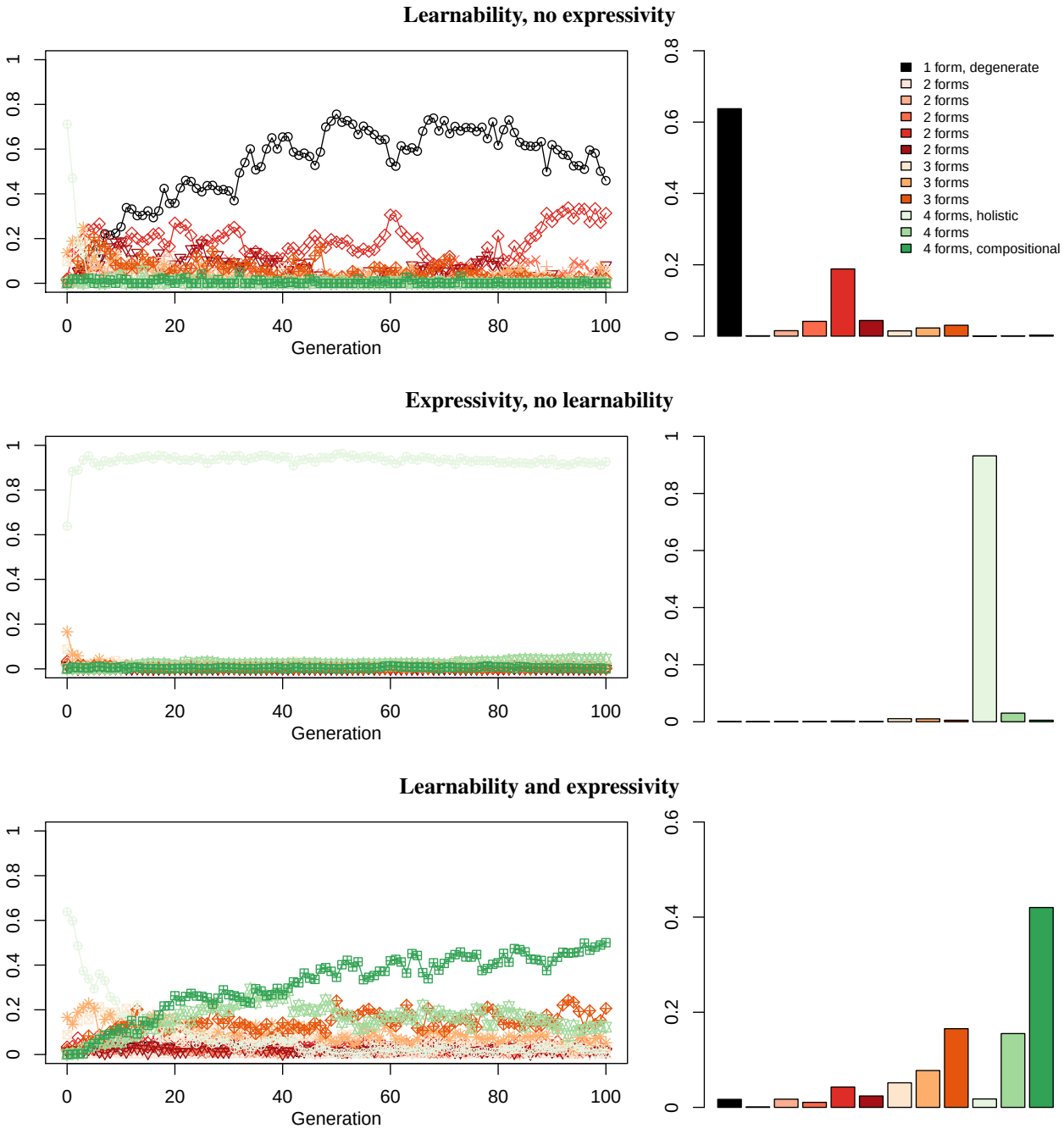

Figure 2: Time courses (left: mean probability of each language class in the last sampled hypothesis of each individual in multiple chains/dyads) and final distributions (right: mean probability as in time courses, averaged over the final 50 generations of those same simulation runs). When learnability is the only pressure (top: average of 20 simulation runs), degenerate languages dominate the final distribution. When expressivity is the only pressure (middle: 50 runs), the original expressive but holistic languages are preserved. When both pressures are at play (bottom: 50 runs), expressive but compositionally-structured languages emerge.

cisely these species, whose communication system is under cultural selection to be learnable but expressive. This is entirely consistent with the predictions of our model, although we would suggest that the expressivity pressures inherent in communication in these species must be rather different from the expressivity pressure in language, with a focus on signalling e.g. individual quality, rather than communicating propositions.

Conclusions

The results from our model support the hypothesis drawn from our review of the modelling and experimental literature on the evolution of communication systems: structure emerges when a system of communication is under pressure to be both expressive (due to communicative interaction) and simple (due to domain-general preferences for compressibility imposed during language learning). Crucially, both these pressures must be in play: pressure for expressivity or simplicity alone does not lead to structure. Linguistic structure therefore can and should be explained as a consequence of cultural evolution: structure in language is a linguistic adaptation, not a biological adaptation, and it is an adaptation in response to competing pressures for expressivity and learnability inherent in language transmission and use.

References

Brighton, H., Kirby, S., & Smith, K. (2005). Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable. In M. Tallerman (Ed.), *Language origins: Perspectives on evolution* (pp. 291–309). Oxford: Oxford University Press.

Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, *2*, 177–226.

Burkett, D., & Griffiths, T. L. (2010). Iterated learning of multiple languages from multiple teachers. In A. D. M. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *The Evolution of Language: Proceedings of the 8th International Conference* (p. 58-65). Singapore: Word Scientific.

Chater, N., & Vitanyi, P. (2003). Simplicity: a unifying principle in cognitive science. *Trends in Cognitive Science*, *7*, 19-22.

Christiansen, M., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*, 489–509.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, *1*, 209-230.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, *4*(3), 258–267.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from? *Cognitive Science*, *31*(6), 961–987.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 173–203). Cambridge: Cambridge University Press.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *PNAS*, *105*(31), 10681-10686.

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *PNAS*, *104*, 5241–5245.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707–784.

Smith, K. (2004). The evolution of vocabulary. *Journal of Theoretical Biology*, *228*(1), 127–142.

Steels, L. (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, *103*, 133–156.

Steels, L. (2012). Self-organization and linguistic selection in language evolution. In L. Steels (Ed.), *Experiments in cultural language evolution.* Amsterdam: John Benjamins.

Theisen-White, C., Kirby, S., & Oberlander, J. (2011). Integrating the horizontal and vertical cultural transmission of novel communication systems. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 956–961). Austin, TX: Cognitive Science Society.

Appendix

For a given set of forms F (where members of the set are either complete forms, e.g. aa, or a partially-specified form, e.g. $a*$, indicating a string-initial a) and a set of meanings M, we can count the number of mappings in a language for which forms from F is associated with meanings from M: we denote this quantity $n(M,F)$. For instance, $n(\{00,10,11,10\},aa) = 4$ for the degenerate language in Table 1, since the form aa is associated with all 4 of these meanings, but 1 for the compositional and holistic languages; $n(0*,a*) = 2$ for the degenerate and compositional languages but 1 for the holistic language, since there is only a single mapping where meaning-initial 0 maps to form-initial a. Our base probability for language l characterised is then:

$$G_0(l) \propto P(l,\mathcal{M},\mathcal{F}).\prod_{m\in\{0*,1*\}} P(l,m,\{a*,b*\}).\prod_{m\in\{*0,*1\}} P(l,m,\{*a,*b\})$$

where we normalise over all possible languages and $P(l,M,F)$ is the prior from (Kirby et al., 2007),

$$P(l,M,F) = \frac{\Gamma(|F|\alpha)}{\Gamma(\alpha)^{|F|}\Gamma(m+|F|\alpha)} \prod_{f\in F} \Gamma(n(M,f)+\alpha),$$

where $\Gamma(x) = (x-1)!$ and m is the number of meanings from $\mathcal{M}$ that unify with M. The parameters σ determines the strength of the preference for simplicity: low σ (we use $\sigma = 1$) strengthens the preference for more compressible languages, higher σ leads to a weaker preference for such languages.